FACULTÉ DE GOUVERNANCE,
SCIENCES ECONOMIQUES ET SOCIALES

Module : Principes d'Inférence Statistique

1. Présentation du module

2. Échantillonnage

Prof. Aziz Ajbilou

1. Présentation générale du module

1. Statistique descriptive et probabilités : des prérequis pour l'inférence statistique ;
2. Inférence statistique : des techniques statistiques fondées sur l'observation ;
3. Objectifs et contenu du module ;
4. Quelques références de base ;
5. Évaluation
6. Travail en groupe

2. Echantillonnage

Statistique descriptive et lois de probabilités : des prérequis pour l'inférence statistique

1

Statistique descriptive :

- Organiser ;
- Résumer ;
- Analyser ;
- interpréter.

Calcul des mesures (synthétiques):

- **de tendance centrale : moyenne, mode, médiane, etc.**
- **de dispersion : variance ou écart type, coefficients de variation, etc.**
- **Coefficient de corrélation**

Données concernent :

- ✓ Une population ;
- ✓ Une sous population

2

Théorie des probabilités :

- Caractère probabiliste
- Caractère Aléatoire

Différentes lois caractérisant une population

- **Notions d'expérience, de probabilité, variabl aléatoire, etc.,**
- **Calcul d'espérance mathématique,**
- **Lois discrètes et lois continues**

Caractérisent une :

- Une population ;
- Une sous population

3

Inférence statistique :

- **Généraliser les résultats d'un échantillon à l'ensemble de la population**

Des techniques statistiques

- **Échantillonnage ;**
- **Tests statistiques**
- **Etc.**

Fondée sur :

- **L'observation à partir d'une sous population**

Statistique descriptive et lois de probabilités : des prérequis pour l'inférence statistique

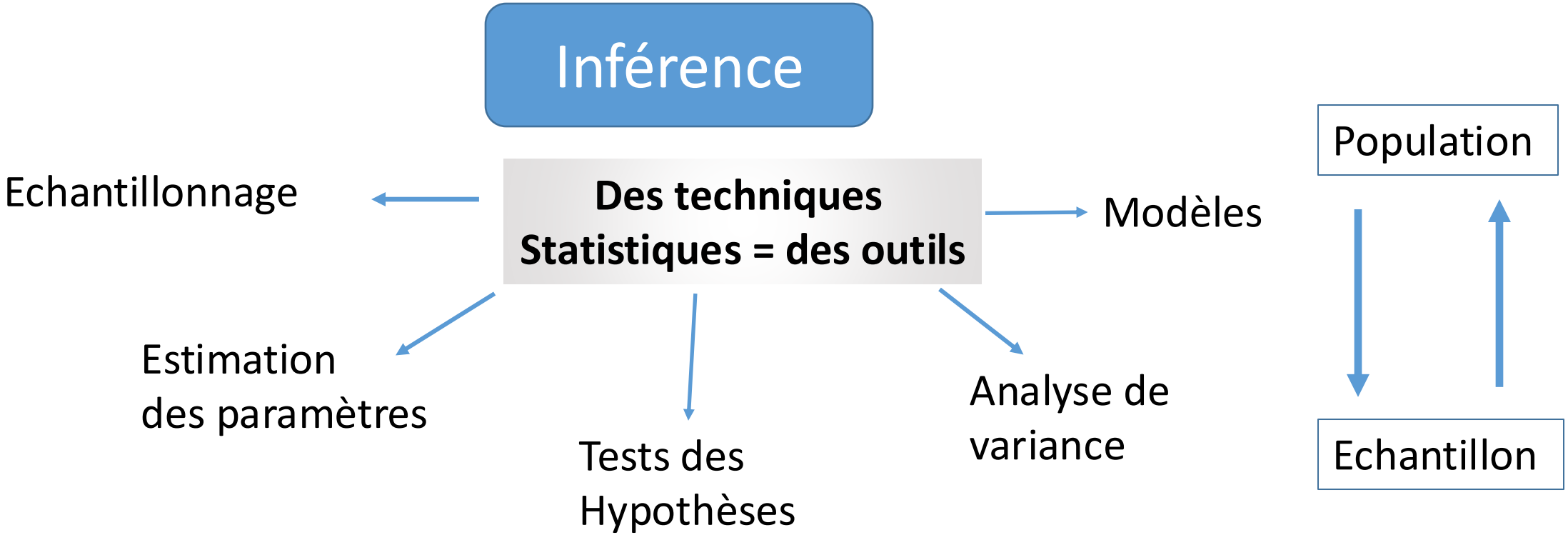
Trois volets essentiels (à titre indicatif)

- **Maitriser le calcul des:**
  - **probabilités ;**
  - **mesures de tendance centrale notamment : la moyenne ou la proportion d'une variable. Le calcul de l'espérance mathématique ;**
  - **mesures de dispersion notamment : la variance, l'écart type.**

- Les distributions de probabilités notamment :
  - **La loi normale ;**
  - **La loi Khi deux ;**
  - **La loi Student ;**
  - **La loi Fisher**
- Savoir interpréter les résultats

L'inférence statistique : des techniques statistiques fondées sur l'observation

Inférence statistique : c'est le fait de généraliser ou d'extrapoler, à l'ensemble de la population étudiée les résultats obtenus à partir d'un échantillon et ce en utilisant des techniques statistiques.



Sur la base de quoi ?:

- des observations ;
- des lois de probabilités

Objectifs et contenu du module

- **Acquérir les bases d'analyse des données statistiques dans ses applications les plus fréquentes en statistique**
  1. Acquérir **le vocabulaire** propre à l'inférence statistique ;
  2. Comprendre **l'utilité des méthodes statistiques** en tant qu' **outil d'aide à la décision**, de **collecte d'information**, de **construction de modèle**, **d'évaluation d'impact**, etc. ;
  3. **Interpréter correctement** des données collectées et des résultats.
- Comprendre les techniques d'inférence statistique et les conditions de leur application
  1. Concepts et Techniques de **tirage d'un échantillon** ;
  2. **Echantillon représentatif** et échantillon **non représentatif** ;
  3. Techniques d'estimation ponctuelle et par intervalle
  4. Tests d'hypothèses et **précise de décision** ;
- Comprendre les techniques d'analyse de variance (comparaison des moyennes)
- Maitriser les **outils de prévision** :
  1. Analyse de régression linéaire ;
  2. Analyse de corrélation ;

Objectifs et Contenus du module

- Présentation du module ;
- Échantillonnage ;
  - ✓ Méthodes de tirage de l'échantillon, etc.
    - Méthodes probabilistes : tirage aléatoire simple, tirage systématique, sondage stratifié, etc.
    - Méthodes non probabilistes
  - ✓ Loi normale Générale et Standard
  - ✓ Loi de Student
  - ✓ Distribution de l'échantillonnage ;
  - ✓ Comment déterminer la taille de l'échantillon ?

- Estimation des paramètres d'une population
  - ✓ Des notions de base ;
  - ✓ Estimation ponctuelle VS estimation par intervalle
  - ✓ Estimation ponctuelle et par intervalle de la moyenne et d'une proportion d'une population

Objectifs et Contenus du module

- Tests d'hypothèses (un seul échantillon ; deux échantillons)
  - ✓ Pourquoi les tests d'hypothèses ?
  - ✓ Formulation des hypothèses à tester ;
  - ✓ Méthodes et étapes d'un test d'hypothèses ;
  - ✓ Interprétation des résultats des tests d'hypothèse, etc.

- Analyse de la variance ;
  - ✓ Objectifs
  - ✓ Conditions d'application ;
  - ✓ Etapes d'une analyse de variance ;
  - ✓ Interprétation des résultats

- Régression linéaire ;
  - ✓ Introduction sur les relations entre variables
  - ✓ Régression simple ;
  - ✓ Régression multiple ;

Quelques références de base

- Douglas A. Lind, William G. Marchal, Samuel A. Wathen. (2012), « Statistical Techniques in Business & Economics », McGraw-Hill in Montréal, 15th ed.
- Anderson, D.R, Sweeney D.J., Williams Th. A. (2011), « Statistics for Business and Economics », South Western.
- **Ken Black (2010), Business Statistics For Contemporary Decision Making, 6th edition, University of Houston—Clear Lake, USA.**
- Gilbert SAPORTA, (2006). Probabilités, analyse des données et statistique. 2ème édition révisée et augmentée. Paris : Editions Technip, 622 pages.
- Douglas. A lind… {et al.], (2006). Méthodes statistiques pour les sciences de la gestion ; adaptation françaises, Sylvie Chomé, Denis Loracque, Roch Ouellet by Chinelière/.

- Site **WWW.hcp.ma ;**

Quelques références de base

Référence principale : Ken Black (2010), Business Statistics For Contemporary Decision Making, 6th edition, University of Houston—Clear Lake, USA

4 Probability 92

UNIT II **DISTRIBUTIONS AND SAMPLING**

5 Discrete Distributions 136
6 Continuous Distributions 178
7 Sampling and Sampling Distributions 216

UNIT III **MAKING INFERENCES ABOUT POPULATION PARAMETERS**

8 Statistical Inference: Estimation for Single Populations 250
9 Statistical Inference: Hypothesis Testing for Single Populations 288
10 Statistical Inferences About Two Populations 342
11 Analysis of Variance and Design of Experiments 402

UNIT IV **REGRESSION ANALYSIS AND FORECASTING**

12 Simple Regression Analysis and Correlation 464
13 Multiple Regression Analysis 516
14 Building Multiple Regression Models 546

Côté logiciel

- Excel
  - ✓ Un lien pour toutes les fonctions statistiques sur Excel avec leur explication : **https://www.excel-pratique.com/fr/index_des_fonctions/statistiques**

- **R**

Présentation générale du cours : évaluation

Evaluation

La note finale est calculée sur la base de :

- 30 % pour l'Examen final ;
- 30 % pour Examen de mi- parcours ;
- 25 % pour le travail en groupe ;
- 15 % pour les Quiz (3): la meilleure note des trois sera comptée

Des travaux à mener

1. Formation des groupes ;
2. Choix d'un sujet d'enquête
3. Appliquer les différents chapitres :
   - Echantillonnage ;
   - Estimation des paramètres ;
   - Tests des hypothèses ;
   - Analyse de la variance
   - Régression simple et multiple

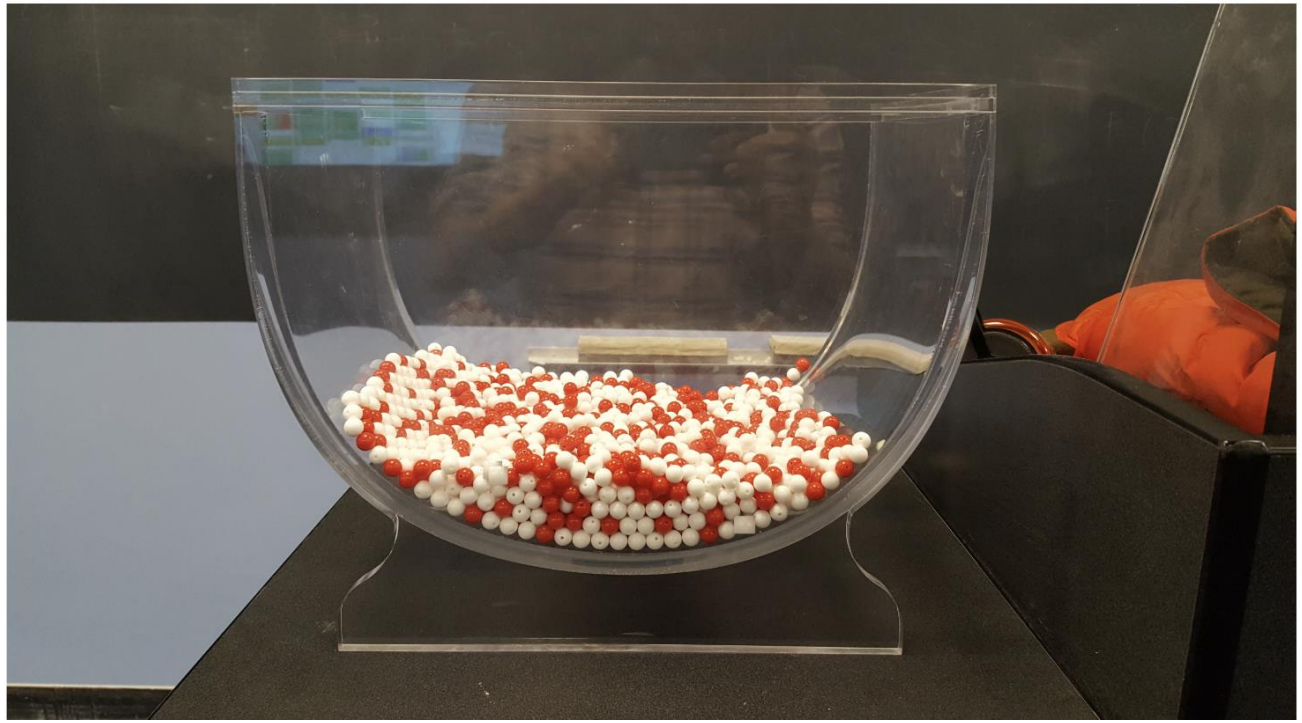
Voir consignes sur Canvas

2. Echantillonnage

1. Eléments introductifs : quelques concepts de base ;

2. Méthodes de tirage d’un échantillon ;
   - Méthodes probabilistes ;
   - Méthodes non probabilistes ;

Exemple



- Quelle est la proportion des boules rouges ?
- Comment procéder ?

1. Eléments introductifs : quelques concepts de base

Population

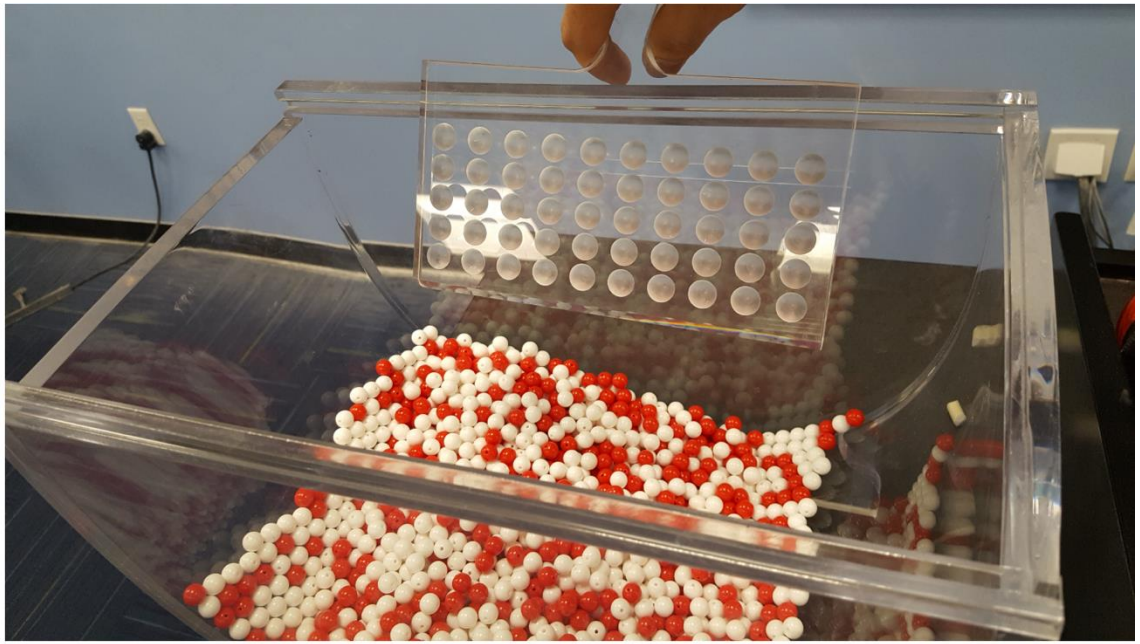
- Un ensemble fini ou non d'éléments ou individus bien identifiés qui ont des caractéristiques, des propriétés communes.
- Il peut s'agir des êtres humains (adultes, enfants, jeunes, personnes âgées, chômeurs, salariés, etc.), d'animaux ou encore d'objets (entreprises, voitures, ordinateurs, etc.).

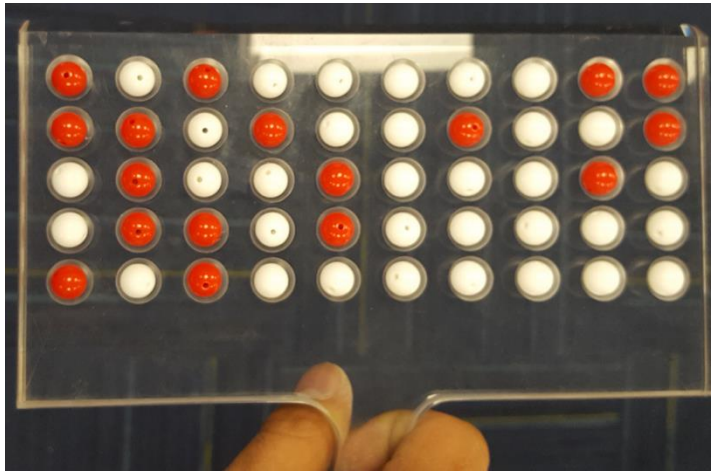
  - ❖ Etudiants de la faculté de la Gouvernance et Sciences Economiques et Sociales ;
  - ❖ Les jeunes diplômés à Rabat en 2020 ;
  - ❖ les chauffeurs de taxi à Tanger ;
  - ❖ Les utilisateurs des smartphones ;
  - ❖ Les utilisateurs d''internet ;
  - ❖ La production d'une entreprise ;
  - ❖ Etc.

1. Eléments introductifs : quelques concepts de base

Echantillon

- Un sous-ensemble d'une population. Cet échantillon doit posséder les mêmes caractéristiques statistiques de la population dont il est issu. Il devrait être représentatif de la population étudiée.





1. Eléments introductifs : quelques concepts de base

Echantillonnage : c'est quoi ?

L'échantillonnage :

- est une procédure de tirage d'un échantillon.
- est largement utilisé dans différents domaines et ce pour collecter des informations utiles sur une population.

le passage de la population à l'échantillon
(du général au particulier)

1. Eléments introductifs : quelques concepts de base

Echantillonnage : pourquoi ?

a. Optimisation du coût de collecte et de traitement des données (généralement les ressources financières sont limitées d'où le recours à un échantillon au lieu à un recensement) ;
b. L'enquête par sondage est plus rapide, surtout lorsque la caractéristique étudiée présente des modifications au cours du temps ;
c. Dans certains cas le recours à l'enquête s'impose vue la nature de la population étudiée et les objectifs visés. C'est le cas lorsque :
    i. L'observation présente un caractère répétitif dans le temps (chaque jour, chaque semaine, chaque mois, etc.) : **l'enquête sur les prix ;**
    ii. Le décideur demande des données détaillées et qui ne peuvent pas faire l'objet d'un recensement (**Enquête niveau de vie, Enquête démographique, Enquête sur l'emploi**) ;
    iii. Si l'accès à la population est difficile ou même impossible, l'échantillon est la seule option.

1. Eléments introductifs : quelques concepts de base

Unité statistique

- La population objet de l'étude est composée d'*individus* (unité de base à laquelle on s'intéresse) appelés aussi *observations* ou *unités statistiques*.

- Chaque individu est supposé « *se démarquer* » sans ambiguïté des autres par une information claire et précise (identifiant "i").

- L'unité statistique peut être :
  - ✓ *simple :*
    - ❖ personne ;
    - ❖ employé dans une entreprise ;
  - ✓ *complexe :*
    - ❖ Ménage ;
    - ❖ Logement ;
    - ❖ Entreprise ;
    - ❖ Zone géographique.

1. Eléments introductifs : quelques concepts de base

Variables ou caractéristiques

- Les individus sont soumis à l'observation à propos d'un certain nombre de caractéristiques d'intérêt : des variables « X » ;

    - X_1 = Age
    - X_2 = Niveau d'instruction
    - X_3 = Nombre d'années d'études ;
    - X_4 = Milieu de résidence
    - X_5 = nombre de personne dans le ménage
    - …..

- Les variables peuvent être :
    - ✓ Quantitatives : Valeurs sont quantitatives
        - ❖ nombre d'années d'études,
        - ❖ chiffre d'affaires,
        - ❖ revenu,… ;
    - ✓ Qualitatives : valeurs sont des modalités, ou catégories, exprimées sous forme littérale ou par un codage numérique :
        - ❖ Diplôme ;
        - ❖ Niveau d'instruction ;
        - ❖ Milieu de résidence ;
        - ❖ …

1. Eléments introductifs : quelques concepts de base

Exemples de variables

- Economie /finance : consommation d'un bien; dépenses, niveau de vie, accès aux crédits…
- Démographie : décès, naissance, migration,…
- Logement : type de logement; ancienneté de logement..
- Social : loisirs, utilisation d'internet, niveau d'instruction, conditions de vie,…
- Politique : intentions de vote, opinions sur une action gouvernementale,…

Pour une entreprise

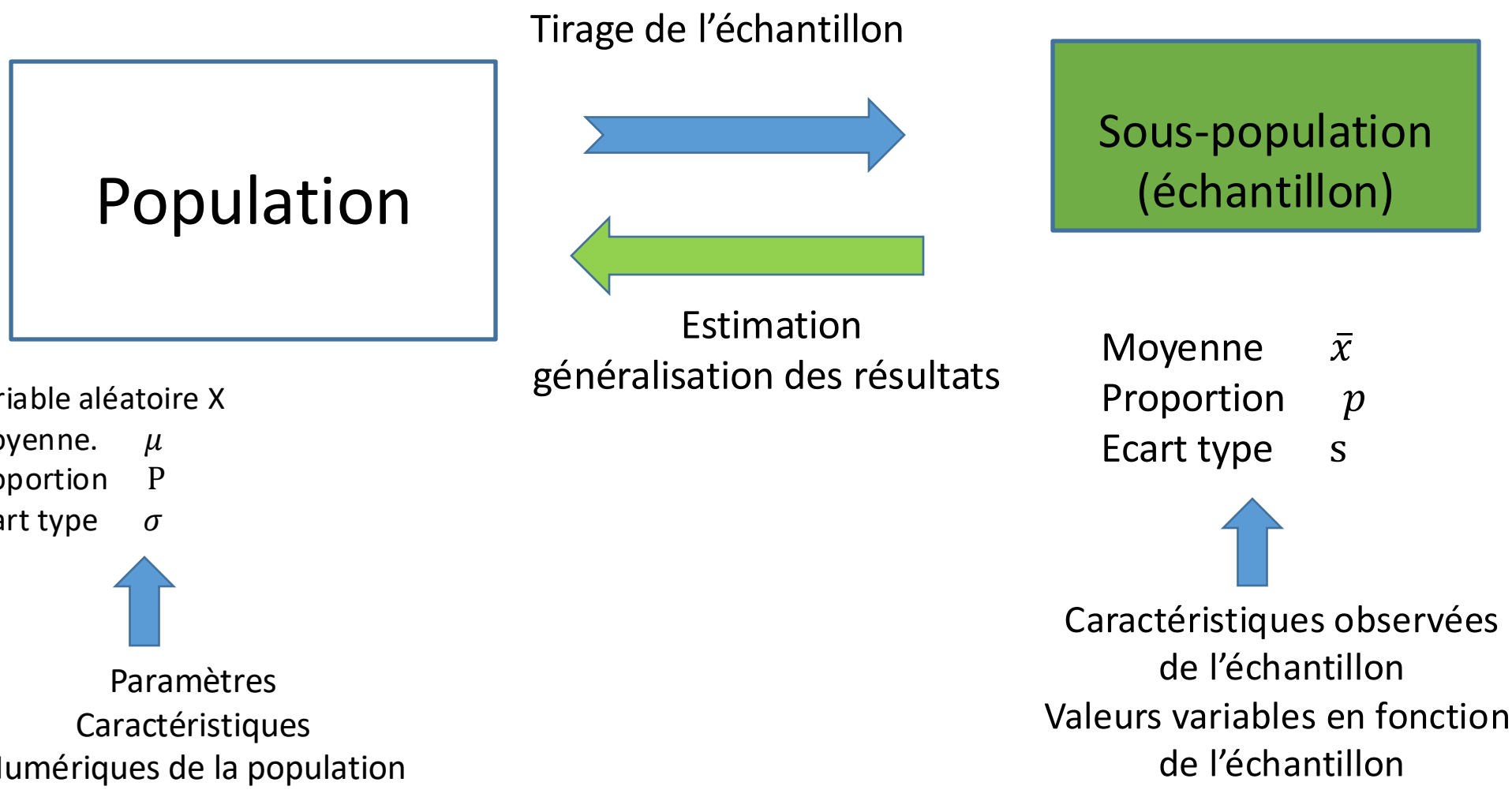
- Type d'activité ;
- Taille de l'entreprise ;
- Chiffre d'affaires ;
- Zone d'activité (quartier industriel/Q.non industriel) ;
- Zone d'activité (U/R)

1. Eléments introductifs : quelques concepts de base

- **Paramètre** :
  - une valeur associée à la variable X au niveau de la population, généralement **inconnue** :
    - revenu moyen de la population;
    - taux de pauvreté de la population;
    - taux de chômage;
    - taux de scolarisation des enfants en âge de scolarité;
    - dépenses moyennes par ménage, etc.)
  - Une valeur fixe (c'est l'information recherchées à travers l'enquête à réaliser)
- **Estimateur :**
  - Statistique calculée à partir **d'un échantillon** qui permet l'estimation d'un paramètre inconnu.
- **Estimation ponctuelle** :
  - Valeur de l'estimateur pour un échantillon observé
  - Dépond de l'échantillon utilisé ;

Paramètre	Valeur fixe (à chercher)	Objet d'intérêt
Estimation	Valeur déterminée à partir de l'échantillon	Un outil

1. Eléments introductifs : quelques concepts de base



1. Eléments introductifs : quelques concepts de base

Observation longitudinale

- On dit que l'observation est **longitudinale** lorsque, pour la **même unité statistique** (individus), **une caractéristique (variable)** est mesurée à **différents moments dans le temps**

  - Permet de tirer des conclusions sur l'évolution temporelle d'une variable ;
    - ✓ Enquêtes démographiques à passages répétés ;
    - ✓ Enquête sur le niveau de vie à passages répétés ;

Observation transversale

- On dit que l'observation est transversale lorsque des données sont collectées auprès des unités statistiques à **un moment donné**.
  - Recensement
  - Enquêtes

Combiner les deux types d'observations

- On peut combiner les deux types d'observation lors d'une même enquête

1. Eléments introductifs : quelques concepts de base

Expérience contrôlée :

- On divise les unités statistiques (individus) en plusieurs groupes (2, 3, 4 …) qui se voient assigner un traitement différent afin d'étudier la validité d'une hypothèse sur l'effet du traitement. Les groupes sont **similaires**, sauf bien sûr au niveau du traitement qui leur est administré. Ceci est nécessaire pour isoler l'effet du traitement.

- Pour isoler l'effet du traitement on distingue :
  - **Groupe de traitement** : groupe soumet au traitement qui fait l'objet de l'étude.
  - **Groupe de contrôle** : groupe sans traitement (agit comme point de référence ou de comparaison).

1. Eléments introductifs : quelques concepts de base

On doit distinguer entre

- La population
  - Un recensement (N)
  - μ , s^2
- L'échantillon
  - Un sondage (n)
  - $\bar{x}$, s^2
- L'échantillonnage
  - Échantillons possibles
  - Moyenne des échantillons ; $\sigma^2(\bar{X})$

1. Eléments introductifs : quelques concepts de base

Base de sondage : c'est quoi ?

Une base de sondage est constituée d'une liste exhaustive des unités statistiques composants la population « P ».

Une base parfaite

1. Possibilité de repérer les unités sans ambiguïté : l'identifiant → liste d'identifiants de bonne qualité
   - **un logement** : commune, l'immeuble et un rang numérique qu'on lui donne dans l'immeuble.
   - un individu : nom et prénom, commune, rue , etc.
2. Exhaustive : autrement on parle d'un défaut de couverture ;
3. Sans double compte ;

→ **Une base de sondage doit être complète, exacte et à jour que possible**

1. Eléments introductifs : quelques concepts de base

Deux types de bases de sondage :

1. Liste : registre d'état civil, des entreprises,, etc.
2. Aréolaire : les unités sont des secteurs géographiques.

Les imperfections d'une base de sondage :

- sous couverture,
- sur couverture,
- répétition,
- classification erronée.

Exemple :

- On veut étudier le parcours **en matière d'emploi** des diplômés de la faculté de Gouvernance et Sciences Economique et Sociale :
    1. Liste exhaustive de tous les étudiants inscrits en 3ème année de la faculté ;
    2. Chaque étudiant est identifié par son matricule ;
    3. Aucun étudiant n'est doublement compté : unicité du matricule.

1. Eléments introductifs : quelques concepts de base

Autres exemples de bases de sondage

1. Base de sondage des entreprises :
   - C'est la liste des entreprises obtenue à partir des sources telles que
     - ✓ le Recensement Economique ;
     - ✓ la Chambre du Commerce et de l'Industrie,…)
2. Base de sondage des exploitations agricoles
   - C'est la liste des exploitations agricoles obtenues à partir :
     - ✓ du Recensement
     - ✓ de la Chambre de l'Agriculture,…
3. Base de sondage des classes des écoles primaires et secondaires
   - C'est la liste de l'ensemble des classes sur la base des informations fournies par le Ministère de l'Education Nationale
4. Base de sondage des structures de santé
   - C'est la liste des structures de santé (Carte sanitaire,) qu'on peut obtenir du ministère de la Santé ;

31

1. Eléments introductifs : quelques concepts de base

Autres exemples de bases de sondage (suite)

- Certaines enquêtes utilisent comme base de sondage des zones géographiques, appelée « base de sondage aréolaire »

  ✓ Cette base aréolaire est obtenue à partir des opérations de cartographie qui précédent le Recensement Général de la Population et de l'Habitat au Maroc (RGPH) ;

  ✓ C'est une liste exhaustive des zones appelées « Districts de Recensement (DR) ». Ces districts de recensement sont caractérisés par des informations auxiliaires à savoir :

    ➢ le code géographique qui précise la région, la province, le cercle et la commune ;

    ➢ Taille du District (nombre de ménages)

32

1. Eléments introductifs : quelques concepts de base

Tirage avec remise

- Un tirage de l'échantillon est dit avec remise si la même unité statistique peut paraître plusieurs fois dans l'échantillon et si l'ordre dans lequel apparaissent les unités statistiques compte.
- Dans ce cas avec une population de taille N et un tirage de l'échantillon de taille n; le nombre d'échantillons possibles serait N^n

Tirage sans remise

- Un tirage est dit sans remise si la même unité statistique ne peut apparaitre qu'une seule foi dans l'échantillon
- le nombre d'échantillons possibles est $C^n_N = \frac{N!}{n!(N-n)!}$

- **En pratique, on utilise le plus souvent les plans de sondage sans remise.**
- **Les deux modes de tirage donnent les mêmes résultats si N est grand**

2. Méthodes de tirage d’un échantillon

2. Méthodes de tirage d’un échantillon

2. Méthodes de tirage d'un échantillon (1)

- On distingue deux groupes de méthodes : Il s'agit des :
  - Méthodes probabilistes
  - Méthodes non probabilistes :

❖Les méthodes probabilistes :

- Se fondent sur la théorie des probabilités : chaque unité de la population a une probabilité d'appartenir à l'échantillon (probabilité d'inclusion ou de sélection). Cette probabilité est connue d'avance.
- Permettent la généralisation des résultats à l'ensemble de la population. Autrement dit, les résultats d'un échantillon fournissent des estimations des paramètres de la population.
- Permettent de calculer l'erreur d'échantillonnage

Des méthodes multiples

- Tirage aléatoire simple
- Tirage aléatoire systématique ;
- Tirage aléatoire stratifié ;
- Tirage aléatoire par grappe ;
- Tirage à plusieurs degrés.

Le but de ces méthodes est d'obtenir un échantillon représentatif de la population étudiée. C'est-à-dire :

- **Une image réduite mais fidèle de la population dont il est extrait. ;**
- **Les caractéristiques de l'échantillon peuvent être utilisées pour estimer celles de la population.**

2. Méthodes de tirage d'un échantillon (2)

❖ Les méthodes non probabilistes ou empiriques ou « à choix raisonné ». Celles-ci :

- ne permettent pas de calculer les probabilités d'inclusion ;
- Introduisent le caractère subjectif dans le choix des unités ;
- ne permettent pas de généraliser les résultats à l'ensemble de la population étudiée ;
- ne permettent pas de calculer l'erreur d'échantillonnage

Des techniques multiples

- Tirage par commodité ;
- Tirage par jugement ;
- Tirage en boule de neige ;
- Tirage par quotas

2.1. Méthodes probabilistes

2.1. Méthodes probabilistes : Tirage aléatoire simple (1)

- Le sondage probabiliste simple est la base de tout sondage probabiliste ;
- Le tirage aléatoire consiste à choisir « n » individus (ou unités statistiques) d'une population de taille « N ».
- Dans ce type de tirage, chaque individu de la population est choisi *au hasard*.
- Chaque individu a la même probabilité de faire partie de l'échantillon.

Tirage aléatoire simple :

- Simple à réaliser ;
- Nécessite la disponibilité d'une base de sondage
- S'applique à des populations homogènes par rapport à la variable étudiée (X) :
  - ✓ les unités sont proches semblables par rapport à X ;

Dans un tirage sans remise

- Le nombre d'échantillons possibles est : $C^n_N = \frac{N!}{n!(N-n)!}$
- La probabilité de tirage d'un individu est : $f = \frac{n}{N}$ appelé également fraction de sondage = taux de sondage

2.1. Méthodes probabilistes : Tirage aléatoire simple (2)

Procédure :

1. Définir clairement la population étudiée;
2. Identifier l'unité statistique
3. Déterminer la base de sondage
4. Numéroter les unités de la base de sondage de 1 à N ;
5. Prendre en considération le nombre de chiffres qui composent (N);
6. Utiliser la table des nombres aléatoires (générés d'une façon aléatoire) ;
7. Choisir un nombre aléatoire d, dont le nombre de chiffres est celui de N ;
8. Se déplacer dans un seul sens jusqu'à obtention de la taille n.

2.1. Méthodes probabilistes : Tirage aléatoire simple (3)

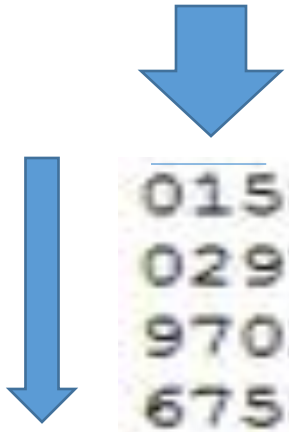


Table des nombres aléatoires

01581	36001	15892	57621	85239
02972	56177	87580	66794	48123
97022	65380	91304	32853	99729
67583	01277	77815	60558	75920
26935	56306	38710	77239	47139
21201	75983	35695	60517	14579
02628	26124	68322	01436	85994
93635	69404	76323	33459	70041
08984	81320	03226	60959	78246
04415	78662	28295	46513	92889
13070	18401	14382	48262	53177
53531	36891	29620	72532	47368
87733	74995	61843	88472	15736
47619	57452	92819	34401	48782
94060	67951	28895	79309	91897

2.1. Méthodes probabilistes : Tirage aléatoire simple (4)

Exemple

1. Soit une population composée de 126 personnes et on veut sélectionner un échantillon de 12 personnes par la méthode de tirage aléatoire simple dont la démarche est la suivante :
   - Les unités de cette population sont numérotées de 1 à 126
   - La taille de la population (N) est composée de 3 chiffres (j=3)
   - On utilise les blocs composés de 3 chiffres (table des nombres aléatoires) ;

2. On recourt à la table des nombres aléatoires et on obtient les unités suivantes (à titre d'exemple) :
   U015 ; U029 ; U026 ; U089 ; U044 ; U017 ; U115;
   U091 ; U003 ; U114 ; U014 ; U099.

01581	36001	15892	57621	85239
02972	56177	87580	66794	48123
97022	65380	91304	32853	99729
67583	01277	77815	60558	75920
26935	56306	38710	77239	47139
21201	75983	35695	60517	14579
02628	26124	68322	01436	85994
93635	69404	76323	33459	70041
08984	81320	03226	60959	78246
04415	78662	28295	46513	92889
13070	18401	14382	48262	53177
53531	36891	29620	72532	47368
87733	74995	61843	88472	15736
47619	57452	92819	34401	48782
94060	67951	28895	79309	91897

2.1. Méthodes probabilistes : Tirage aléatoire systématique (1)

- Population homogène par rapport à v.a. X ;
- Disponibilité de la base de sondage

- Unités statistiques ordonnées de 1 à N ;
- Choix d'une unité de départ d'une manière aléatoire ;
- Choix des autres unités à partir de la première unité statistique selon un intervalle fixe (appelé pas de sondage) jusqu'à parvenir à la taille de l'échantillon désirée.

Procédure de sélection

- Détermination du pas de sondage noté r
  Le pas de sondage est la partie entière du rapport. r= $\frac{N}{n}$
- Choisir une unité de départ d ;
- Les autres unités sont déterminées comme suit : d, d+r, d+2r, d+ 3r, ……. Jusqu'à obtention de la taille n.

2.1. Méthodes probabilistes : Tirage aléatoire systématique (2)

Exemple

- On dispose de la liste de présence de 252 étudiants (ordonnée de 1 à 252) ; N= 252
- On veut choisir 14 étudiants ; n = 14
- On calcule r=N/n = 252/14 = 18 , donc r= 18
- Choix aléatoire du point de départ par exemple « d » = 9 ;
- l'échantillon est composé des unités correspondantes aux rangs suivants :

d=9	C'est l'unité	U9
9+1*18	C'est l'unité	U27
9+ 2*18	C'est l'unité	U45
9+3*18	C'est l'unité	U63
9+4*18	C'est l'unité	U81
9+5*18	C'est l'unité	U99
9+6*18	C'est l'unité	U117

9+7*18	C'est l'unité	U135
9+8*18	C'est l'unité	U153
9+9*18	C'est l'unité	U171
9+10*18	C'est l'unité	U189
9+11*18	C'est l'unité	U208
9+12*18	C'est l'unité	U225
9+13*18	C'est l'unité	U243

2.1. Méthodes probabilistes : Tirage aléatoire stratifié (1)

- Les tirages aléatoires simple et systématique se basent sur l'hypothèse de l'homogénéité de la population par rapport à la variable étudiée (X).

- Cette hypothèse est difficilement réalisable : les populations sont souvent hétérogènes;

- S'il existe dans la base de sondage un critère permettant de distinguer a priori entre les individus, on aura tout à gagner à utiliser cette information pour répartir l'échantillon dans chaque sous-population.

2.1. Méthodes probabilistes : Tirage aléatoire stratifié (1)

- L'idée est de former des groupes relativement homogènes, appelés strates et notés , S_h, h= 1,.., H.
- La Strate S_h : Sous-ensemble homogène de la population possédant une ou plusieurs caractéristiques communes ;
- Les strates sont mutuellement exclusives. Les strates S_h sont de taille « N_h » tel que

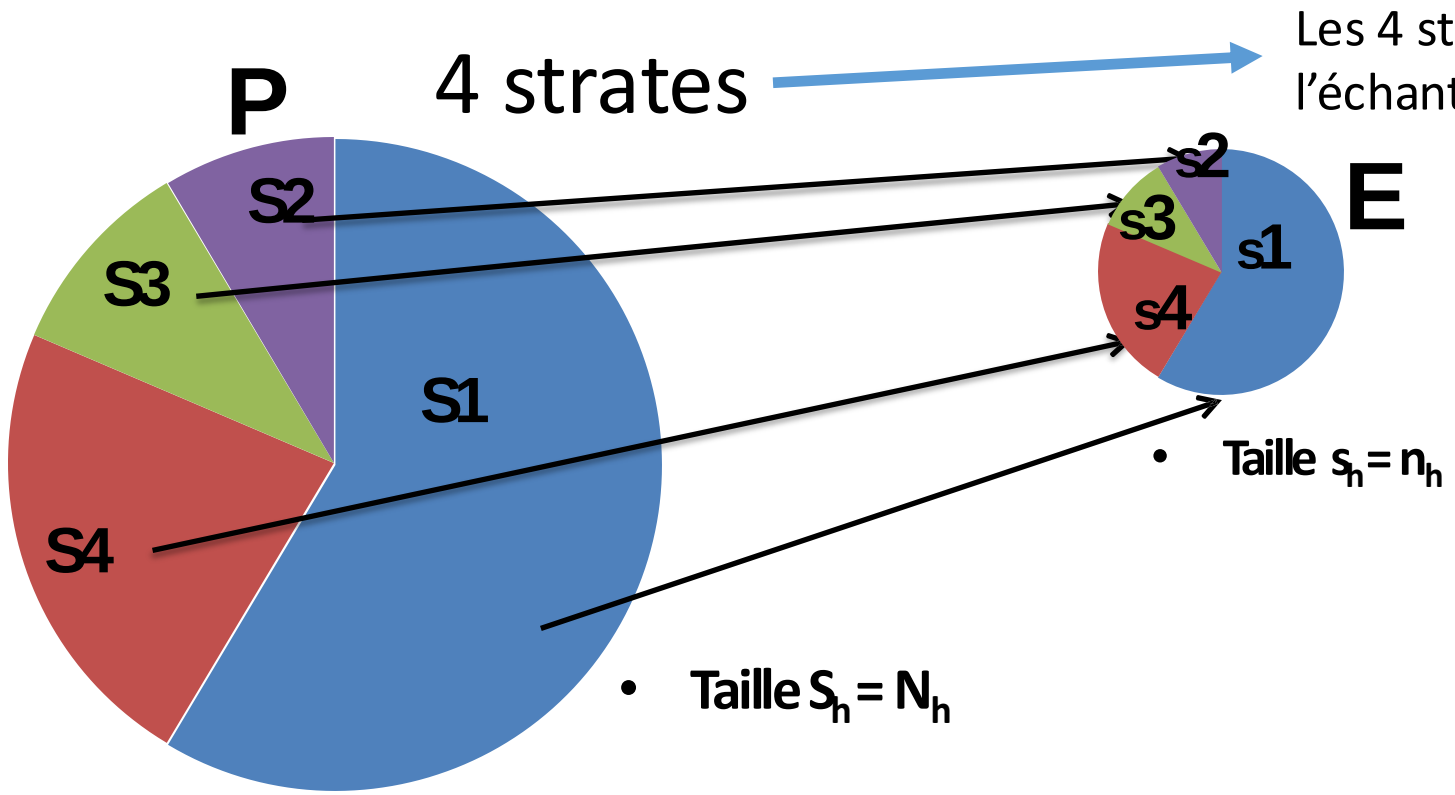
  $\sum N_h = N$
- On effectue au niveau de chaque strate « h » un Tirage aléatoire simple ou un tirage aléatoire systématique d'un échantillon de taille « n_h » individus parmi « N_h » et

  $\sum n_h = n$

Remarque

Toutes les strates qui composent la population sont représentées dans l'échantillon

2.1. Méthodes probabilistes : Tirage aléatoire stratifié (2)



Les 4 strates sont représentées dans l'échantillon

Schéma simplifié d'un Tirage Aléatoire Stratifié avec 4 strates

Choix de la variable de stratification :

- Disponibilité des critères dans la base de sondage ;
- Pertinence des différents critères pour créer des strates homogènes.

Ceci nécessite une connaissance venant d'études réalisées antérieurement.

2.1. Méthodes probabilistes : Tirage aléatoire stratifié (3)

Exemple

- On considère **une population composée**, pour simplifier, de 5 éléments (unités statistiques).
- Les valeurs d'une variable aléatoire dans cette population est :

(13, 15, 17, 25, 30) ;

- La moyenne de cette variable est $\bar{X}$ = (13+15+17+25+30)/5 = 20
- **C'est la moyenne de la population. C'est ce que on cherche à estimer à travers le tirage d'un échantillon.**

2.1. Méthodes probabilistes : Tirage aléatoire stratifié (3)

Exemple (suite)

- On désire tirer un échantillon de taille 2.
- Le tirage utilisé est celui dit tirage aléatoire simple.
- Chaque unité statistique a la même probabilité.
- Le taux de sondage est : 2/5
- Le nombre total d'échantillons possible en utilisant un tirage sans remise est : 10 échantillons possibles.

2.1. Méthodes probabilistes : Tirage aléatoire stratifié (3)

Ces échantillons avec leurs moyennes sont sont :

(13, 15)	moyenne égale à	**14**
(13, 17)	moyenne égale à	**15**
(13, 25)	moyenne égale à	19
(13, 30)	moyenne égale à	21,5
(15, 17)	moyenne égale à	**16**

(15, 25)	moyenne égale à	20
(15, 30)	moyenne égale à	22,5
(17, 25)	moyenne égale à	21
(17, 30)	moyenne égale à	**23,5**
(25, 30)	moyenne égale à	**27,5**

Échantillons extrêmes

- Les moyennes des échantillons extrêmes s'écartent trop de **la moyenne globale = 20**
- pour éviter cette situation on cherche dans la base de sondage un critère permettant de distinguer *a priori les différentes* catégories (en deux sous groupes). On utilise cette information pour répartir l'échantillon dans chaque sous- population.

2.1. Méthodes probabilistes : Tirage aléatoire stratifié (4)

On forme deux strates

- Strate n°1 est composée des unités de couleur Blue {**13 15 17**}
- Strate n°2 est composée des unités de couleur Rouge {**25 30**}

- La variable de stratification est **la couleur des unités**

- On procède au tirage aléatoire simple d'une unité par strate et on obtient les échantillons possibles

		Echantillons possibles					
	Taille de strates	1	2	3	4	5	6
Strate 1	3	13	13	15	15	17	17
Strrate 2	**2**	**25**	**30**	**25**	**30**	**25**	**30**
Moyenne de chanque échantillon	5 unités	17,8	19,8	19	21	20,2	22,2
Moyenne globale		**20**					

$$\bar{x}_1 = \frac{3}{5} * 13 + \frac{2}{5} * 25 = 17{,}8$$
$$\bar{x}_2 = \frac{3}{5} * 13 + \frac{2}{5} * 30 = 19{,}8$$
$$\bar{x}_3 = \frac{3}{5} * 15 + \frac{2}{5} * 25 = 19{,}0$$
$$\bar{x}_4 = \frac{3}{5} * 15 + \frac{2}{5} * 30 = 21{,}0$$
$$\bar{x}_5 = \frac{3}{5} * 17 + \frac{2}{5} * 25 = 20{,}2$$
$$\bar{x}_6 = \frac{3}{5} * 17 + \frac{2}{5} * 30 = 22{,}2$$

2.1. Méthodes probabilistes : Tirage aléatoire stratifié (5)

Cette stratification a permis :

- de réduire le nombre d'échantillons possibles de 10 à 6 ;
- d'éliminer les échantillons extrêmes ;
- D'obtenir des moyennes des échantillons beaucoup plus resserrée autour de la moyenne globale que dans le cas du tirage aléatoire simple : les valeurs extrêmes sont moins éloignées,
- Moins de variabilité entre les moyennes. L'erreur type (c'est-à-dire la racine carrée de la variance des six valeurs) vaut :
  - ✓ 1,40 pour les moyennes avec stratification contre
  - ✓ 3,95 pour les moyennes sans stratification (c'est-à-dire Tirage aléatoire simple)

- On dit que la stratification permet d'augmenter la précision des estimations ou la qualité des estimateurs (à traiter dans le chapitre des estimations)

2.1. Méthodes probabilistes : Tirage aléatoire stratifié (6)

Enquête	Exemples de variables de stratification
Auprès des **ménages**	• Niveau d'instruction du chef de ménage ou la personne de référence ; • Type d'habitat ; milieu de résidence ou région de résidence • Revenu du ménage • Age du chef de ménage ou la personne de référence
Auprès des entreprises	• Taille de l'entreprise ; • Région d'activité ; • Secteur d'activité de l'entreprise ; • Ancienneté de l'entreprise.
ONG	• Région d'activité ; • Type d'activité ;
Des jeunes	• Niveau d'instruction ; • Âge ; • Région.

- **Donnez des exemples d'enquêtes ; précisez l'unité statistique ;**
- **Proposez une variable de stratification pour chaque cas.**

2.1. Méthodes probabilistes : Tirage aléatoire par grappe (1)

Ce type de tirage est indiqué lorsque :

1. La base de sondage des individus est inexistante ou de mauvaise qualité.
   - Il est plus facile de constituer une liste de logements ou d'entreprises que de constituer une liste d'individus ou d'employés (annuaires des entreprises/associations) ;
2. On veut limiter les coûts de collecte de données sur le terrain :
   - Les individus d'une grappe sont généralement voisins ;
   - Pour enquêter un échantillon de 1000 personnes, il suffit de choisir 250 ménages et enquêter 4 personnes par ménage au lieu de 1000.

Comment faire pour constituer les grappes ?

1. On divise la population en sous-ensembles appelés grappes et notés G_h (h=1,…H). Les grappes doivent être les plus hétérogènes possibles.
2. On sélectionne un échantillon de grappes et on procède à l'enquête de **TOUS les individus** constituant l'unité (la grappe) sélectionnée , c'est-à-dire exhaustivement
   - ✓ $n_h=N_h$ pour toute grappe G_h sélectionné
   - ✓ $n_h=0$ pour toute grappe non sélectionnée

2.1. Méthodes probabilistes : Tirage aléatoire par grappe (2)

Des exemples

Exemple 1 : variable d'intérêt X = emploi

1. Grappe d'individus = ménage
2. Unité statistique : personnes en âge d'activités (15 ans et plus)

Exemple 2 : variable d'intérêt X = dépenses alimentaires

1. Grappe de logements = immeuble/ groupe de maisons
2. Unité statistique : ménages et donc les individus composant le ménage

Exemple 3 : variable d'intérêt X = conditions de travail

1. Grappe d'employés/membres = entreprise/association
2. Unité statistique : employés/membres de l'association

- **Donnez des exemples d'enquêtes ; précisez l'unité statistique ;**
- **Donnez des exemples de grappes à former.**

2.1. Méthodes probabilistes : Tirage aléatoire a plusieurs degrés (1)

- C'est un tirage où l'échantillon final est constitué après plusieurs tirages au sort, imbriqués les uns dans les autres.

- **Il peut être un mélange des tirages précédents pour éviter des coûts trop importants ou des conditions qui ne se prêtent pas à l'application de ces méthodes.**

Le principal avantage de ce type de tirage est :

- **de réduire le coût de l'enquête, notamment celui des déplacements.**

- **de permettre dans certaines situations de pallier à l'inexistence d'une base de sondage.**

2.1. Méthodes probabilistes : Tirage aléatoire a plusieurs degrés (2)

- **Exemple 1 : tirage à deux degrés**

  1. Sélection d'un échantillon de villes,
  2. puis sélection des logements à l'intérieur des villes choisies

- **Exemple 2 : tirage à deux degrés**

  1. Sélection d'un échantillon de grandes écoles

  2. Sélection d'un échantillon d'étudiants au sein de chaque grande école choisie